Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 1 (2015), pp. 401-407

http://www.ripublication.com

Forecasting the Dengue Hemorrhagic Fever Cases Using Seasonal ARIMA Model in Chonburi, Thailand

J. Mekparyup[1] S. Kornpetpanee[2] and K. Saithanu[3]

[1,2]College of research methodology and cognitive science, Burapha University

[1,3]Department of Mathematics, Faculty of Science, Burapha University

169 Muang, Chonburi, Thailand

[1]jatupat@buu.ac.th, [2]suchada@buu.ac.th, [3]ksaithan@buu.ac.th

Abstract

The research's objectivity proposed the approach to forecast the dengue hemorrhagic fever (DHF) cases in Chonburi, Thailand using time series technique which was the seasonal ARIMA model. The majority of this research was planning to cope with epidemic of dengue fever to reduce the number of DHF cases infected in the future. The seasonal ARIMA model with parameters defined p=1, d=0, q=2, P=0, D=1 and Q=2 with S=12 and no constant or noted ARIMA(1,0,2)$(0,1,2)_{12}$ was fitted using time series data since 2007 to 2013 with root mean square error (RMSE) of 46.71 and the correlation coefficient (R) between the observed and forecasted DHF cases was 0.912 with p-value=0.000.

Mathematics Subject Classification: 62-07

Keywords: ARIMA model

INTRODUCTION

Dengue hemorrhagic fever (DHF) has become an important global issue when the World Health Organization (WHO) announced that the number of infections increased dramatically 30 times in the last 50 years and expected that a significant number of patients with DHF of 50 to 100 million cases. It shows that the nearly half world population is at risk of dengue virus infection [1]. For areas of outbreaks, there are more than 100 countries, mostly in the tropical and subtropical regions [2][3][4]. In Thailand, DHF is a major problem for public health and medicine because the DHF patients trend to increase rapidly each year [5]. Chonburi is one of the eastern province in Thailand facing the epidemic of DHF [6] so any protective strategy coped with dengue fever outbreak should be planed rushly. Then the objective of this research is to forecast the number of DHF cases in 2014 and 2015 to help the authorized unit such as the Bureau of Epidemiology to project reducing the epidemic of DHF in Chonburi.

MATERIALS AND METHODS

The number of DHF cases in Chonburi was collected from the Bureau of Epidemiology, National Trustworthy and Competent Authority in Epidemiological Surveillance and Investigation, Thailand since January 2007 to December 2013.

1. THE STATIONARY OF TIME SERIES

Before fitting the seasonal ARIMA model, it is remarked that the time series assumed to be stationary. Stationary of time series can be validated by Autocorrelation function (ACF) and Partial autocorrelation function (PACF). Autocorrelation is an association between Z_t and Z_{t-k} but the parameters are unknown so the correlation can be calculated by Sample autocorrelation function (SACF) as following Equation 1.

$$r_k = \frac{\sum_{t=1}^{n-1} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{n} (Z_t - \bar{Z})^2} \qquad (1)$$

where Z_t be a variable at time t, k be time lag and $\bar{Z} = \sum_{t=1}^{n} Z_t / n$. Time series is stationary when the graph of SACF declines rapidly to zero or the graph is zero all the time. Box and Jenkins proposed 50 the number of the observed values should be taken at least 50 [7]. Similarly, Partial autocorrelation is an association between $(Z_t - \hat{Z}_t)$ and $(Z_{t-k} - \hat{Z}_{t-k})$ which can be determined by Sample partial autocorrelation function (SPACF) as well as following Equation 2.

$$\hat{\phi}_{kk} = \begin{cases} r_1; \ k = 1 \\ \dfrac{r_k - \sum_{j=1}^{k-1} \hat{\phi}_{(k-1)j} r_{k-1}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{(k-1)j} r_j}; \ k = 2, 3, \ldots \end{cases} \qquad (2)$$

where $\hat{\phi}_{kj} = \hat{\phi}_{(k-1)j} - \hat{\phi}_{kk} \hat{\phi}_{(k-1)(k-j)}$; $j = 1, 2, \ldots, k-1$. Determining stationary condition of time series for SPACF is the same as consideration of the SACF graph.

2. THE SEASONAL ARIMA MODEL

After the stationary condition of time series verified, multiplicative seasonal autoregressive integrated moving average model of order $(p,d,q)(P,D,Q)_S$ or ARIMA$(p,d,q)(P,D,Q)_S$ is determined by parameter estimations, p, d, q, P, D and Q, as following Equation 3.

$$\begin{aligned}&(1-\phi_1 B-\ldots-\phi_p B^p)(1-\Phi_1 B^s-\ldots-\Phi_P B^{P_s})(1-B)^d(1-L)^D Z_t \\ &=(1-\theta_1 B-\ldots-\theta_q B^q)(1-\Theta_1 B^s-\ldots-\Theta_Q B^{Q_s})a_t\end{aligned} \quad (3)$$

where a_t be the random error at time t, ϕ_i and Φ_I be regression coefficient at time i and I; i=1, 2, …, p; I=1, 2, …, P, θ_j and Θ_J be moving average coefficient at time j and J;j=1, 2, …, q; J=1, 2, …, Q. There are many methods to estimate all these parameters, p, d, q, P, D and Q, such as moment method, conditional least square method, maximum likelihood method [7], etc. Then the model is fitted if the forecast error is within $(1-\alpha)100\%$ confidence interval by checking the residuals of the SACF graph.

3. THE LJUNG-BOX Q STATISTIC

The next step is to validate selected model using the Ljung-Box Q statistic as Equation 4 [8] when the parameters are estimated in Equation 3.

$$Q=\{(n-d)[(n-d)+2]\}\sum_{j=1}^{k}\frac{r_j^2}{[(n-d)-j]} \quad (4)$$

where k be the distance of time lag, n be the number of observations of the time series, d be the order of the time series variances and r_j be autocorrelation at time lag j. The Q statistic is chi-square distribution with degrees of freedom $k-p-q$ when n is large.

4. THE STATISTICAL MEASURES OF THE GOODNESS OF FIT

In order to validate of the seasonal ARIMA model, the root mean square error (RMSE) and the correlation coefficient (R) are calculated consequently as Equation 5 and 6.

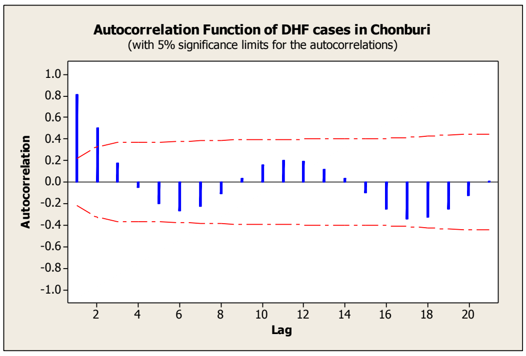
$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(OBS_i-FOR_i)^2} \quad (5)$$

$$R = \frac{\sum_{i=1}^{n} (OBS_i - \overline{OBS})(FOR_i - \overline{FOR})}{\sqrt{\sum_{i=1}^{n} (OBS_i - \overline{OBS})^2 \sum_{i=1}^{n} (FOR_i - \overline{FOR})^2}} \tag{6}$$

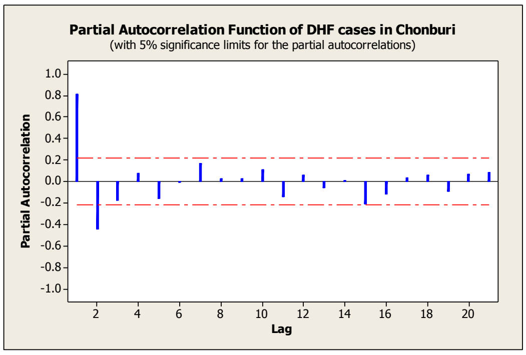
where OBS_i be the observation data, $\overline{OBS}$ be the average of OBS, FOR_i be the forecasted data and $\overline{FOR}$ be the average of FOR.

RESULTS AND DISCUSSION

The SACF and SPACF graph were plotted to testify the stationary validation of DHF cases in Chonburi for the parameters estimation of the seasonal ARIMA model as Figure 1a and 1b consequently.



(a)



(b)

Figure 1: Checking the stationary condition of the DHF cases in Chonburi; (a) SACF graph, (b) SPACF graph

As Figure 1 and 2 showed that the time series was nonstationary then DHF cases were transformed by the difference. Finally, the parameters of the seasonal ARIMA model were estimated by p=1, d=0, q=2, P=0, D=1 and Q=2 with S=12 and no constant or ARIMA(1,0,2) $(0,1,2)_{12}$ shown in Equation 7 and the coefficient values of the model were shown in Table 1.

$$(1-\phi_1 B^1)(1-B^{12})Z_t = (1-\theta_1 B^1 - \theta_2 B^2)(1-\Theta_1 B^{12} - \Theta_2 B^{24})a_t$$

or

$$Z_t = \phi_1 Z_{t-1} + Z_{t-12} - \phi_1 Z_{t-13} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \Theta_1 a_{t-12}$$

$$+\theta_1 a_{t-13} + \theta_2 \Theta_1 a_{t-15} - \Theta_2 a_{t-24} + \theta_1 \Theta_2 a_{t-25} + \theta_2 \Theta_2 a_{t-26} \tag{7}$$

Table 1: The estimator values and testing of the parameters of ARIMA(1,0,2) $(0,1,2)_{12}$

Parameters	Estimation	Error	T statistic	p-value
ϕ_1(AR 1)	0.4135	0.1604	2.58	0.012
θ_1(MA 1)	–0.8002	0.1366	–5.86	0.000
θ_2(MA 2)	–0.5425	0.1196	–4.54	0.000
Θ_1(SMA 12)	1.2972	0.1200	10.81	0.000
Θ_2(SMA 24)	–0.5180	0.1490	–3.48	0.001

As of Table 1, all parameters of the model were tested following; ϕ_1(T=2.58, p-value=0.012), θ_1(T = –5.86, p-value=0.000), θ_2(T = –4.54, p-value=0.000), Θ_1(T=10.81, p-value=0.000) and Θ_2(T = –3.48, p-value=0.001), so all parameters were not equal to zero. Then the seasonal model to forecast the DHF cases in Chonburi was written as of Equation 8.

$$Z_t = 0.4135Z_{t-1} + Z_{t-12} - 0.4135Z_{t-13} + a_t - 0.4135a_{t-1} + 0.5425a_{t-2}$$
$$-1.2972a_{t-12} - 0.8002a_{t-13} - 0.7037a_{t-15} + 0.5180a_{t-24} + 0.4145a_{t-25}$$
$$+0.2810a_{t-26} \quad (8)$$

The result of validating the fitted model was displayed in Figure 2 indicated that the forecast errors were within 95% confidence interval.

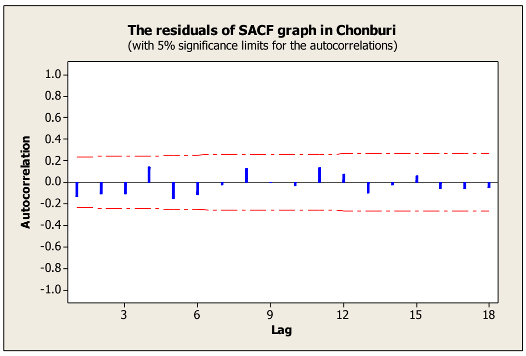


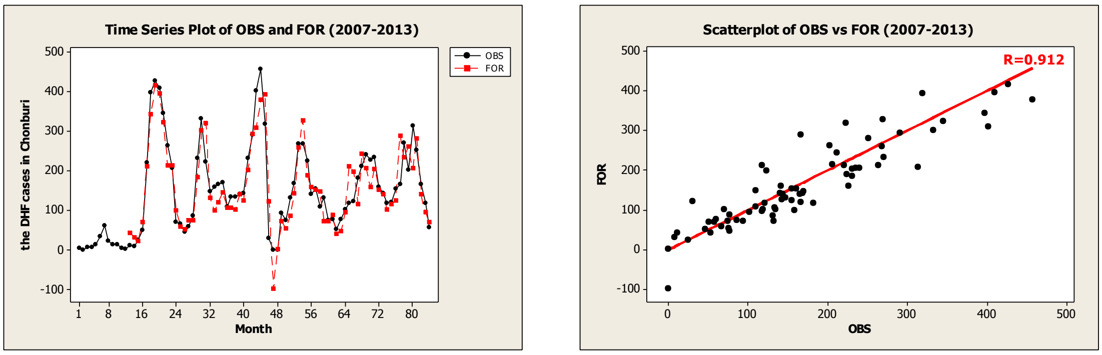
Figure 2: The residuals of SACF graph of ARIMA(1,0,2) $(0,1,2)_{12}$

The Ljung-Box Q statistic was generated to verify the forecasted DHF cases in next 4 years illustrated in Table 2. The model, ARIMA(1,0,2) $(0,1,2)_{12}$, was appropriate to forecast the DHF cases in Chonburi with the RMSE=46.71 by Q_{12}=11.1 (p-value=0.136), Q_{24}=21.4 (p-value=0.313), Q_{36}=28.5 (p-value=0.597) and Q_{48}=34.1 (p-value=0.832).

Table 2: The Q statistic of ARIMA(1,0,2) $(0,1,2)_{12}$

Lag	12	24	36	48
Chi-square statistic	11.1	21.4	28.5	34.1
Degrees of freedom	7	19	31	43
p-value	0.136	0.313	0.597	0.832

Comparison between the observed and forecasted DHF cases was demonstrated consequently by time series plot and scatter plot as of Figure 3a and 3b. The correlation coefficient of 0.912 (p-value=0.000) disclosed that the forecasted data was very close to the observed data. The ARIMA(1,0,2) $(0,1,2)_{12}$ model was then fitted to forecast the DHF cases in Chonburi year 2014 and 2015 presented in Table 3.



(a) (b)

Figure 3: Comparison between the OBS and the FOR; (a) Time series plot, (b) Scatter plot

Table 3: The forecasted DHF cases in Chonburi using ARIMA(1,0,2) $(0,1,2)_{12}$

Year \ Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2014	82.11	57.13	86.36	103.58	232.19	279.47	307.66	244.01	215.67	114.68	117.29	68.71
2015	98.3	85.56	108.47	135.03	195.05	240.77	241.43	257.25	232.92	151.87	145.8	89.75

ACKNOWLEDGEMENT

The authors wish to thank to the Bureau of Epidemiology, National Trustworthy and Competent Authority in Epidemiological Surveillance and Investigation, Thailand for kindly support providing the data.

REFERENCES

[1] World Health Organization, 2013, “Dengue Control,” Retrieved July 30, 2013, from World Health Organization Web site: http://www.who.int/denguecontrol/en

[2] Rasgon, J. L., 2011, “Dengue fever: Mosquitoes attacked from within,” Nature, 476(7361), 407-408.

[3] Rajapakse, S., Rodrigo, C., & Rajapakse, A., 2012, “Treatment of dengue fever,” Infection and drug resistance, 5, 103-112.

[4] Brady, O. J., Gething, P. W., Bhatt, S., Messina, J. P., Brownstein, J. S., Hoen, A. G., Moyes, C. L., Farlow, A. W., Scott, T. W., & Hay, S. I., 2012, “Refining the global spatial limits of dengue virus transmission by evidence-based consensus,” PLoS neglected tropical diseases, 6(8), e1760.

[5] the Bureau of Epidemiology, 2013, “Dengue fever,” Retrived July 30, 2013, from the Bureau of Epidemiology Web site: http://www.boe.moph.go.th/fact/Dengue_ Haemorrhagic_Fever.htm

[6] Pattamak, C., 2014, “Prediction of the number of the cases of dengue hemorrhagic fever in Eastern region of Thailand, 2013-2017,” Journal of Health Science, 23, 804-812.

[7] Box, G. E. P., & Jenkins, G. M., 1976, “Time series analysis forecasting and control (2nd ed.),” San Francisco: Holden-day.

[8] Ljung, G. M., & Box, G. E., 1979, “The likelihood function of stationary autoregressive-moving average models,” Biometrika, 66(2), 265-270.